

Mapping value sensitive design onto AI for social good principles

Year: 2021 | Citations: 179 | Authors: Steven Umbrello, I. van de Poel

Abstract

Value sensitive design (VSD) is an established method for integrating values into technical design. It has been applied to different technologies and, more recently, to artificial intelligence (AI). We argue that AI poses a number of challenges specific to VSD that require a somewhat modified VSD approach. Machine learning (ML), in particular, poses two challenges. First, humans may not understand how an AI system learns certain things. This requires paying attention to values such as transparency, explicability, and accountability. Second, ML may lead to AI systems adapting in ways that 'disembody' the values embedded in them. To address this, we propose a threefold modified VSD approach: (1) integrating a known set of VSD principles (AI4SG) as design norms from which more specific design requirements can be derived; (2) distinguishing between values that are promoted and respected by the design to ensure outcomes that not only do no harm but also contribute to good, and (3) extending the VSD process to encompass the whole life cycle of an AI technology to monitor unintended value consequences and redesign as needed. We illustrate our VSD for AI approach with an example use case of a SARS-CoV-2 contact tracing app.